2021-01

Technology Trust: System Information Impact on Autonomous Systems Adoption in High-Risk Applications

Anderson, Michael G.; Mun, Johnathan C.

ARJ

Anderson, Michael G., and Johnathan C. Mun. "Technology Trust: System Information Impact on Autonomous Systems Adoption in High-Risk Applications." Defense Acquisition Research Journal: A Publication of the Defense Acquisition University 28.1 (2021).

http://hdl.handle.net/10945/67435

TECHNOLOGY TRUST: SYSTEM INFORMATION IMPACT ON AUTONOMOUS SYSTEMS ADOPTION IN HIGH-RISK APPLICATIONS

Michael G. Anderson and Johnathan C. Mun

As autonomous systems become more capable, end users must make decisions about how and when to deploy such technology. The use and adoption of a technology to replace a human actor depends on its ability to perform a desired task and on the user's experience-based trust that it will do so. The development of experience-based trust in autonomous systems is costly, and it carries a high risk of physical harm to operators. This work focuses on identifying a methodology for technology discovery that reduces the need for experience-based trust and contributes to increased adoption of autonomous systems. The main research hypothesis is that manipulating the presentation of technical information can influence the initial formation of trust by functioning as a surrogate for experience-based trust, and that trust in technology can be captured through an anthropomorphic hierarchy of system attributes.

DOI: *https:// doi.org/10.22594/10.22594/dau.19-841.28.01*
Keywords: *Technology Trust, Autonomous Systems, Technology Risk Metrics, Anthropomorphic Hierarchy, Technology System Attributes*

Image designed by Nicole Brate

The use of technology by the Department of Defense (DoD) depends on its ability to perform a desired task. Many issues associated with trust in technology are increasing in importance as the U.S. military begins to acquire and deploy autonomous systems. To ensure the effective adoption of new innovations in technology, researchers need to establish a system of metrics that justify a level of technology trust. This article has the explicit goal of investigating and recommending trust metrics by applying advanced analytical methodologies to increase the speed and effectiveness of the adoption of new technologies. This investigation proceeds by participating in an evaluation of technologies for use in evolving, high-risk military applications. The trust metrics are measured in terms of the technology acceptance versus system control.

Technology Trust

The 2016 Defense Science Board report on autonomy (David & Nielsen, 2016) identifies trust as central to DoD's success in the broader adoption of autonomy. This article studies the potential for introducing trust metrics on the evaluation and selection of technologies. The work participates in an ongoing assessment of autonomous systems for use in high-risk military applications throughout fiscal year 2019. A model is developed that optimizes the cognitive impacts of these trust metrics as they relate to the technology selection and adoption process. The approach will be extensible and can be adopted into private industry.

> **"This article has the explicit goal of investigating and recommending trust metrics by applying advanced analytical methodologies to increase the speed and effectiveness of the adoption of new technologies.**

Research Problem

The recent increase in the use and deployment of sophisticated technologies by other countries is a disruptive threat to the United States' technological superiority. The rapidly changing technology landscape requires DoD laboratories to increase the speed at which they adopt new technologies (David & Nielsen, 2016). With declining budgets in research, it is imperative that the DoD establish new methods for rapidly adopting and effectively deploying new and emerging technologies whenever possible. The goal of this article is to establish and measure a comprehensive trust metric for individual components of technologies, such as autonomous systems used in high-risk military applications. The development of a trust metric serves two purposes: first, as a surrogate for experience-based trust by contributing to the formation of initial-trust and, second, as a collection tool for capturing experience-based trust data.

This work emerges from the general question, "Can humans develop trust in complex systems without direct experience and a complete understanding of the technology?" Theories in anthropomorphism (assigned human attributes to technology) and system hierarchy hold promise in their ability to reduce complexity and improve the acceptance of complex systems. Thus, the specific research question posited by this article is "How does system information affect the adoption of autonomous systems used in high-risk military applications?"

To that end, this study attempts to answer the following questions:

1. How does the anthropomorphic categorization and presentation of technology affect the development of trust in technologies used in high-risk military applications? The constructs researched include:
    - Hardware
    - Algorithms
    - Links

2. How do varying levels of system control affect the development of trust in technologies used in high-risk military applications? The constructs researched include:
   - Perceived ease of use
   - Perceived usefulness
   - Intent to use
3. Does a causal relationship exist between an anthropomorphic hierarchy of system information and the acceptance of autonomous systems?

Literature Review

This article was initiated through informal interviews that attempted to identify the factors that contribute to the use of technology in high-risk environments. The participants were a small group of military personnel who have deployed with technology that posed great risk of physical harm should it fail. A majority of this group experienced significant injury due to the failure of technology, and the potential for bias was noted. A series of open-ended questions were provided to discuss what the users did or did not like about using technology in high-risk scenarios. The initial coding of interviews revealed the following three exploratory research themes:

1. Hands-on experience with technology is critical for establishing trust, and a team-based reputation for a technology is as important as personal experience.
2. Personal investment in a mission is key to learning and accepting new and complex technology.
3. Users operating in high-risk environments favor simple technology containing only the features needed to accomplish a mission and may reject new and complex technology in favor of older and more trusted systems.

These themes all have implications for the adoption of autonomous systems within the DoD. Advanced robotic systems have the ability to improve performance in a number of military roles while reducing risk to humans, and it is important to understand how to improve the adoption of such systems within the DoD. This initial research focused on technology in dangerous environments and reveals that adoption is highly dependent on the ability of the user to obtain the knowledge necessary to develop trust. This theme led to our initial literature review on understanding trust and how it applies to technology adoption.

Trust

Castelfranchi and Falcone (2010) review 72 definitions of what it means to know something well enough to trust, and their work found a great deal of confusion and ambiguity surrounding the use of this term. As a result, a limited unity on a definition of trust is accepted across research disciplines. However, two themes emerged from the many definitions of trust: (a) the basic premise of trust involves two actors, and (b) trust is a relationship in which one entity relies on someone, or something, based on a given criterion.

> **Advanced robotic systems have the ability to improve performance in a number of military roles while reducing risk to humans, and it is important to understand how to improve the adoption of such systems within the DoD.**

Adams and Webb (2002) describe two broad processes of developing trust between two persons. The first is defined as "experience-based trust," which develops through repeated engagements, and the second is called "reason-based trust," which develops in the absence of direct experience.

Rempel et al. (1985) address three factors that influence the development of experience-based trust: competence, benevolence, and integrity. Their work also discusses the significance of the mental motivation behind the desire to establish a relationship and finds it strongly correlated to the factors that influence trust. Their work confirms the second exploratory research theme that emphasizes the importance of personal investment.

Technology

The past research on interpersonal trust applies in many ways to trust in technology. This study examined literature that contributes to the development of a methodology of technology discovery leading to trust in technology. The potential for integrating interpersonal trust research into technology trust was discussed by McKnight et al. (2011). This research found that interpersonal trust is based on a trustor's expectations and reliance on a trustee to perform as expected through benevolence, even though the trustee possesses the volition to choose to do what is right or what is wrong. Because technology does not possess volition (ability to choose), Knight observed, some researchers went as far as to dismiss the idea of trust in technology as irrelevant.

A theory relevant to measuring and characterizing trust is found in the technology acceptance model (TAM) developed by Fred Davis in the late 1980s. This model plays a significant role in the majority of research investigating the factors and attributes that influence the acceptance of a technology. Venkatesh and Bala (2008) present the TAM's ability to predict and measure individual adoption and use of technology. The TAM assesses the behavioral intention to use a technology through two constructs: perceived usefulness (PU), which is defined as the extent to which a person believes that using a technology will enhance his or her job performance; and perceived ease of use (PEOU), which is defined as the degree to which a person believes that using a technology will be free of effort. These two variables are used to establish a relationship between external influences and potential system usage (Gefen et al., 2003).

> **In some military scenarios, developing experience-based trust presents high levels of risk for physical injury and harm.**

Tétard and Collan (2009) address the challenges of adopting new technology for high-risk scenarios in their work on the lazy-user, also called efficient-user theory. This theory states that users select the technology that demands the least amount of effort to do the job. The application of this theory places technology users at a disadvantage, particularly in high-risk military applications where our exploratory research indicates that users are known to avoid more capable technology for systems that are easier to understand. If an experience-based proxy can improve the accuracy of developing trust through increased technology literacy, it may lead to increased acceptance of more complex and capable technologies, thereby reducing the influence of the efficient-user theory. This leads to our third theme identified in exploratory research, "Users operating in high-risk environments favor simple technology containing only the features needed to accomplish a mission and may reject new and complex technology in favor of older and more trusted systems."

Experimental Design

The previous section discussed how a "trust-discovery" methodology could contribute to improved understanding of how people develop trust in machines. This understanding could lead to the development of a technology-literate workforce capable of accurately assessing new technology for a given operational scenario. The literature review strongly suggests that the manipulation of system information may influence technology trust.

This experiment investigates the formation of trust in technology and how it influences the adoption of autonomous systems for use in high-risk military applications. The formation of trust in technology is governed by two constructs: reason-based trust and experience-based trust. Existing literature presents the case for increased accuracy in technology selection through the development of experience-based trust. However, the development of experience-based trust is financially burdensome and takes much longer to form than reason-based trust. In some military scenarios, developing experience-based trust presents high levels of risk for physical injury and harm.

Experiment Introduction

This experiment is designed to research the manipulation of system information and study any influence on the formation of reason-based trust in autonomous systems used in high-risk military applications. The desired outcome of this work is the identification of causal relationships between system attributes and technology acceptance that can replace some of the burden required to develop experience-based trust. In other words, can a reason-based trust method be used to replace experience-based methods?

The experiment is designed in two-phases. Phase one is a group-administered experimental survey that employs manipulations of multiple theories of system information and technology acceptance to collect data on reason-based trust in systems with varying levels of system control. Phase two consists of administering the same survey, following extensive field testing and experimentation of the phase one systems, to collect data on experience-based trust. Trust is measured as an "intent to use" and based on responses to the TAM.

Anthropomorphism

The complexity of modern technology makes it difficult to establish generalizable categories capable of capturing system information and functioning as a proxy for experience-based trust. One area of research relevant to the establishment of technology categories involves anthropomorphism—the attribution of human traits to nonhuman entities to increase a trustor's ability to understand and accept complex technology.

Schaefer et al. (2016) and Waytz et al. (2014) identified anthropomorphism as a system factor that contributes to the development of human trust for robots. Reported cases in Pak et al. (2012) examine where the tendency to anthropomorphize technology leads to situations in which humans give a higher degree of trust to a technology than is warranted. The inverse of this situation also exists in the development of a lack of trust in a human teammate caused by the introduction of technology with more capability and reliability. In this experiment, anthropomorphism is assessed for its ability to influence technology trust.

System Hierarchy

In this work, we hypothesize that statistically significant differences will result in technology trust by introducing system information through an anthropomorphic hierarchy. Over a period of 10 years, the authors of this article provided instruction to third-year university engineering students on the topics of digital design and computer architecture. The predominant challenge reported by students in end-of-year course evaluations was difficulty synthesizing the highly complex components of a computer into a usable system. Based on student feedback, an anthropomorphic hierarchy was developed to structure the components of computer architecture to a more familiar format. This hierarchy provided students with the context needed to understand how the pieces of a computer function together to create a whole system. The work resulted in improved student ability to describe a computer from the elemental circuits up to the most advanced concepts of computer engineering such as compilers and operating systems.

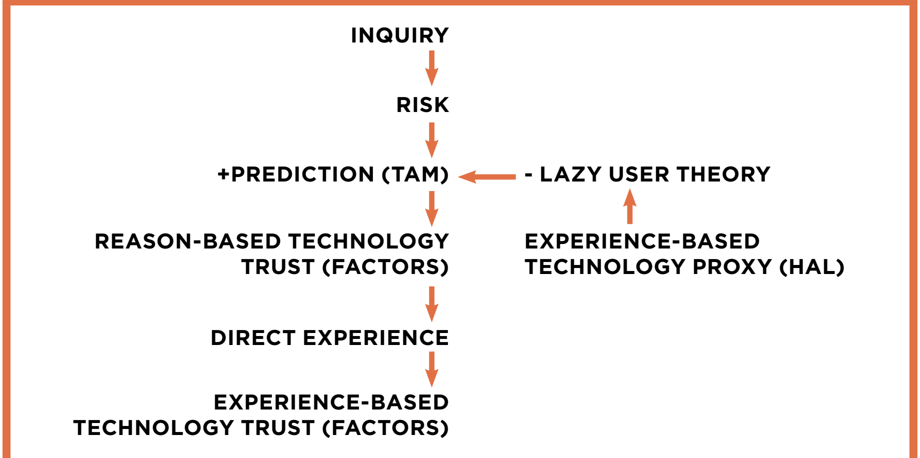
FIGURE 1. ANTHROPOMORPHIC TECHNOLOGY HIERARCHY

HARDWARE Processor Circuits	⟷	**BODY** Brain Nervous System
ALGORITHMS Data OS	⟷	**THOUGHT** Knowledge Wisdom
LINKS Inputs Outputs	⟷	**SENSES** Sight, Smell, Hear, Taste, Touch

To establish an invariant system hierarchy for use in measuring both reason-based and experience-based technology trust, we introduce the anthropomorphic categories of hardware, algorithms, and links (HAL) as illustrated in Figure 1.

To increase the value of this hierarchy, we further conceptualized a HAL score of trustworthiness. The values of each HAL metric are proposed to range from 0 to 100, and lead to an equally weighted maximum score (indicating most trustworthy) of 300. Future research is needed to identify the weights for the HAL score to accurately reflect the overall impact on trust. Since field experimentation has not been conducted, we introduce the HAL categories in the experiment without any associated "score." The HAL hierarchy is used to organize system information and provide a framework for future experience-based trust proxy research as shown in Figure 2.

FIGURE 2. HAL SCORE EXPERIMENTAL MODEL



The Experiment

The experiment was conducted using factorial design. Two independent variables are used: system information and level of system control. The level of system information is varied between two conditions: Less Information (Less Info) and More Information (More Info). The Less Info condition presented system information using vendor-provided datasheets. The More Info condition introduced the same system information but carefully organized under the HAL hierarchy. The system control is varied between three levels: Direct, Remote, and Autonomous. Two dependent variables are used: (a) level of risk associated with the loss of a system attribute, and (b) trust, as measured by the TAM.

Procedures

This article uses data provided by an ongoing external experiment (Appendix A). Phase one of that experiment is a group-administered survey that employs manipulations of system presentation. The phase one surveys were administered to a military unit responsible for the operational assessment of the three technologies (Direct-Control, Remote-Control, Autonomous). Each of the two phase-one surveys (Less Info, More Info) was completed by 20–25 subjects. Demographics such as age, military specialty, and exposure to similar technologies were captured to assess internal validity. Phase two consists of administering the same survey following extensive field testing by 15 subjects from the same military unit tasked in phase one. The phase two results are captured to provide external validity for the phase one results.

> **One area of research relevant to the establishment of technology categories involves anthropomorphism—the attribution of human traits to nonhuman entities to increase a trustor's ability to understand and accept complex technology.**

Phase one of the experiment was conducted in a controlled and distraction-free classroom environment and involved the participation of two randomly selected groups of active-duty military tasked with a new high-risk mission. The two groups participated in separate morning sessions lasting 1 hour each. The start time for the second session was immediately following completion of the first session. Both groups were provided with identical overviews of a high-risk military scenario that would be completed by deploying three technology systems rather than human operators. The independent variable "system presentation" was manipulated between the first and second groups as Less Info and More Info. The second independent variable—"system control"—is provided to each participant in both groups in the form of the three different technologies. Appendix B lists the details of the survey questions as well as variable names and codes that are presented in the next section.

Results

Data Are Only Somewhat Normal

The research questions we are attempting to answer in this section follow.

- *Are the data considered sufficiently normal?*
- *Can we apply conventional parametric methods, or do we need more advanced nonparametric methods to analyze the data?*

Tables 1 and 2 show the results from randomly selected variables. The results are mixed. This means that although these are statistically significant in some areas, they may not be practically significant enough to justify normality.

TABLE 1. VAR1 NORMALITY TESTS

Best-Fitting Distributions: VAR1					
Rank	**Akaike**	**Anderson**	**Kolmogorov**	**Kuiper's**	**Schwartz**
1	Cosine	**Normal**	GenPareto	**Normal**	Cosine
2	Lognml3Arith	Logistic	Weibull	Logistic	Lognml3Arith
3	Weibull	TDist	GumbelMin	TDist	Weibull
4	**Normal**	Weibull	Triangular	Cosine	**Normal**
5	Gamma	GumbelMax	**Normal**	Weibull	Gamma
MAPE %					
1	19.0136%	19.0915%	N/A	19.4214%	19.0136%
2	19.3421%	19.2969%	19.5824%	19.4214%	19.3421%
3	19.3665%	19.4732%	24.8250%	19.4370%	19.3665%
4	19.4297%	20.0214%	21.2316%	19.4732%	19.4297%
5	19.4575%	21.8529%	19.6539%	19.6312%	19.4575%

Best Fit Rank : 5
Fit Name : Normal
Kolmogorov-Smirnov Statistic : 0.153350
Mean : 3.721371
Sigma : 1.250896
p value : **0.614791**
Actual to Theoretical Four Moments :
3.739130 1.053884 -0.190064 -1.168769;
3.721371 1.250896 0.000000 0.000000;

Nonparametric Shapiro-Wilk Test for Normality
(Royston Algorithm)
Shapiro-Wilks : 0.865946
SW *P*-value : **0.005368**
Null hypothesis: The data are normally distributed

Note. MAPE = Mean Absolute Percentage Error; VAR = Variable

TABLE 2. VAR105 NORMALITY TESTS

Best-Fitting Distributions: VAR105					
Rank	**Akaike**	**Anderson**	**Kolmogorov**	**Kuiper's**	**Schwartz**
1	Cosine	TDist	Weibull	TDist	Cosine
2	Uniform	Gamma	Uniform	GumbelMax	Uniform
3	Triangular	Normal	GumbelMax	Weibull	Triangular
4	Weibull	GumbelMin	LognmlArith	Laplace	Weibull
5	TDist	Logistic	**Normal**	GumbelMin	TDist
MAPE %					
1	20.2105%	20.4966%	25.4875%	20.4966%	20.2105%
2	20.3731%	21.5868%	19.8248%	21.8717%	20.3731%
3	20.4260%	22.5328%	25.6700%	22.2282%	20.4260%
4	20.4405%	22.6221%	23.5800%	23.3391%	20.4405%
5	20.4966%	22.9440%	20.7503%	24.0731%	20.4966%

Best Fit Rank : 5
Fit Name : Normal
Kolmogorov-Smirnov Statistic : 0.175000
Mean : 3.647780
Sigma : 1.105387
p value : **0.531299**
Actual to Theoretical Four Moments :
3.550000 1.050063 -0.146220 -1.072526;
3.647780 1.105387 0.000000 0.000000;

Nonparametric Shapiro-Wilk Test for Normality
(Royston Algorithm)
Shapiro-Wilks : 0.880332
SW *P*-value : **0.017937**
Null hypothesis: The data are normally distributed

We conclude that:

- The survey data are only somewhat normally distributed under certain circumstances, and we cannot state complete normality to fully justify standard modeling approaches.
- The data are ordinal and quasi-interval, with limited truncation between 1 and 5, and are limited to between 19 and 23 observations.
- Both parametric and nonparametric methods will be used, and this mixed approach is therefore justified.

Therefore, going forward, both parametric and nonparametric tests will be conducted whenever appropriate, and their results will be compared for corroboration.

Hotelling's T-Squared Distribution in Statistics

The research questions we are attempting to answer in this section are:

- *When all the survey responses for each subgroup are taken together as a whole, are there statistical differences in the responses?*
  - *Are the perceptions of the Direct-Control system different when Less Info is provided, More Info is available, or a hands-on experiment is conducted?*
  - *Are the perceptions of the Remote-Control system different when Less Info is provided, More Info is available, or a hands-on experiment is conducted?*
  - *Are the perceptions of the Autonomous system different when Less Info is provided, More Info is available, or a hands-on experiment is conducted?*

Tables 3 and 4 show a sampling of the results from the Hotelling T^2 test. The null hypothesis is that no statistical differences result from using a parametric Hotelling T^2 test, where all of the survey responses in each of the subcategories, when taken together, simultaneously do not indicate that any differences are discernible between the two groups tested.

TABLE 3. HOTELLING

Hotelling Test Groups	*P*-value	Variables Tested
Less Info vs. More Info	0.5863	VAR1:VAR14 vs. VAR101:VAR114
Less Info vs. More Info	0.7998	VAR15:VAR25 vs. VAR115:VAR125
Less Info vs. More Info	0.3515	VAR26:VAR36 vs. VAR126:VAR136
Less Info vs. More Info	0.2084	VAR37:VAR47 vs. VAR137:VAR147
Less Info vs. More Info	0.7095	VAR48:VAR51 vs. VAR148:VAR151
Less Info vs. More Info	0.4475	VAR52:VAR54 vs. VAR152:VAR154
Less Info vs. Experiment	0.0000	VAR15:VAR25 vs. VAR415:VAR425
Less Info vs. Experiment	0.0144	VAR26:VAR36 vs. VAR426:VAR436
Less Info vs. Experiment	0.0793	VAR37:VAR47 vs. VAR437:VAR447
More Info vs. Experiment	0.0000	VAR115:VAR125 vs. VAR415:VAR425
More Info vs. Experiment	0.1215	VAR126:VAR136 vs. VAR426:VAR436
More Info vs. Experiment	0.3232	VAR137:VAR147 vs. VAR437:VAR447

TABLE 4. HOTELLING FOR GROUP A6 VS. GROUP B6

VAR52; VAR53; VAR54 vs. VAR152; VAR153; VAR154
D1, D2, D3 vs. D1, D2, D3
Hotelling T-Square: Two Independent Variable
Equal Variance with Multiple Related Measures

Hotelling T2 2.85372
F Statistic 0.90484
P-value 0.44753
Null hypothesis tested is that there is zero difference between all the related variables compared across the two groups.

Covariance GROUP 1			
	VAR52	**VAR53**	**VAR54**
VAR52	0.00000	0.00000	0.00000
VAR53	0.00000	15.56621	13.61660
VAR54	0.00000	13.61660	17.73419
Covariance GROUP 2			
VAR152	0.23947	-0.26711	-1.00395
VAR153	-0.26711	9.35461	10.39408
VAR154	-1.00395	10.39408	15.11776
Covariance POOLED			
VAR152	0.11098	-0.12378	-0.46524
VAR153	-0.12378	12.68766	12.12324
VAR154	-0.46524	12.12324	16.52170

We conclude that:

- The results indicate that no perceivable differences exist between the Less Info and More Info groups in the Pre-experiment stage (comparing all subelements of group A to all subelements of group B).
- When comparing the Less Info Pre-experiment group against the Post-experiment group, we see a statistically significant difference among the responses. The trend seems to be that more difference is shown between group A (Less Info) and group C (Post-experiment) than between group B (More Info) and group C.
- In addition, the significance is higher for Direct-Control systems than Remote-Control systems, which in turn, is more significant than Autonomous systems.

Bonferroni Test

The research question we are attempting to answer in this section is:

- *When all the survey responses for each subgroup are taken individually, are there statistical differences in the responses?*

Table 5 shows a sampling of the results from the Bonferroni test. While the previous parametric Hotelling test looks at all subcategories in each group compared to all the subcategories in the second group, the parametric Bonferroni test compares one pair of the subgroups at a time, like the *t*-test. The difference is the Bonferroni test accounts for the added degrees of freedom with multiple simultaneous pairwise tests.

TABLE 5. BONFERRONI TEST

Simultaneous Confidence Intervals
Mean Difference of Null is 0
Model Inputs:

	VAR48; VAR 148; C1	**VAR49; VAR 149; C2**	**VAR50; VAR 150; C3**	**VAR51; VAR 151; C4**
Mean Difference	0.0522	-0.3283	-0.0152	-0.0457
Variance Group 1	1.6917	1.5336	0.8024	0.5850
Variance Group 2	1.4105	0.5553	0.3658	0.5553
Pooled Variance	1.2496	1.0393	0.7746	0.7558
F-Critical	2.6190	2.6190	2.6190	2.6190
T-Critical	3.3620	3.3620	3.3620	3.3620
Standard Error	0.3820	0.3178	0.2368	0.2311
Lower Confidence	-1.2323	-1.3966	-0.8115	-0.8225
Upper Confidence	1.3366	0.7401	0.7810	0.7312
Within Confidence?	Yes	Yes	Yes	Yes
Bonferroni Critical	2.6127	2.6127	2.6127	2.6127
Lower Confidence	-1.4760	-1.5993	-0.9626	-0.9700
Upper Confidence	1.5803	0.9428	0.9321	0.8786
Within Confidence?	Yes	Yes	Yes	Yes

Null hypothesis: The individual expected differences are equal to zero.

We conclude that:

- In all the tests, we did not detect any statistical significance, and find that all subgroups are statistically identical. This implies that additional testing is required.

The Three Systems Are Perceived Differently

The research question we are attempting to answer in this section is:

- *Are the three systems statistically different in their main characteristics?*

Forty-three separate Single Variable Multiple Treatment ANOVA models were run. Table 6 shows the statistically significant results from the ANOVA models. Out of the 43 models, 21 show statistical significance. ANOVA tests each of the survey questions in each of the three systems independently. For example, when testing VAR20, VAR31, VAR42, we see that at least one or more of these three variables are statistically different from one another.

TABLE 6. ANOVA I

ANOVA	*P*-value
VAR20; VAR31; VAR42	0.0008
VAR21; VAR32; VAR43	0.0903
VAR120; VAR131; VAR142	0.0264
VAR124; VAR135; VAR146	0.0362
VAR229; VAR240; VAR251	0.0000
VAR230; VAR241; VAR252	0.0000
AR231; VAR242; VAR253	0.0000
VAR232; VAR243; VAR254	0.0002
VAR233; VAR244; VAR255	0.0601
VAR237; VAR248; VAR259	0.0004
VAR238; VAR249; VAR260	0.0000
VAR239; VAR250; VAR261	0.0000
VAR266; VAR276; VAR286	0.0000
VAR267; VAR277; VAR287	0.0285
VAR268; VAR278; VAR288	0.0003
VAR269; VAR279; VAR289	0.0020
VAR270; VAR280; VAR290	0.0000
VAR271; VAR281; VAR291	0.0000
VAR272; VAR282; VAR292	0.0351
VAR273; VAR283; VAR293	0.0002
VAR274; VAR284; VAR294	0.0000

We conclude that:

- The ANOVA results support the results from the Hotelling T^2 tests, where we see that group A is statistically significantly different than group B and group C; and group B is statistically significantly different than group C.

The ANOVA test looks at the individual questions within each of these groups to identify which questions returned different responses for each of the systems in the three different testing environments (Pre-experiment less data, Pre-experiment more data, and Post-experiment).

The Three Surveys Provide New Significantly Valuable Information

The research question we are attempting to answer in this section is:

- *Do the added information and hands-on experimentation provide additional value-added insights?*

Thirty-three separate Single Variable Multiple Treatment ANOVA models were also run to test the individual questions among the three systems among the three groups (i.e., for each of the survey questions, if each of the three systems has similarities or differences among the Pre-experiment Less Info, Pre-experiment More Info, and Post-experiment groups). Table 7 shows the statistically significant results from the ANOVA models. Out of the 33 models, 16 show statistical significance ($\alpha = 0.05$).

TABLE 7. ANOVA II

Model	*P*-value
ANOVA on VAR15; VAR115; VAR415	0.0000
ANOVA on VAR16; VAR116; VAR416	0.0000
ANOVA on VAR17; VAR117; VAR417	0.0000
ANOVA on VAR18; VAR118; VAR418	0.0000
ANOVA on VAR18; VAR118; VAR418	0.0008
ANOVA on VAR20; VAR120; VAR420	0.0003
ANOVA on VAR21; VAR121; VAR421	0.0001
ANOVA on VAR22; VAR122; VAR422	0.0000
ANOVA on VAR23; VAR123; VAR423	0.0001
ANOVA on VAR24; VAR124; VAR424	0.0000
ANOVA on VAR28; VAR128; VAR428	0.0232
ANOVA on VAR29; VAR129; VAR429	0.0157
ANOVA on VAR31; VAR131; VAR431	0.0114
ANOVA on VAR35; VAR135; VAR435	0.0089
ANOVA on VAR38; VAR138; VAR438	0.0472
ANOVA on VAR43; VAR143; VAR443	0.0324

We conclude that:

- Direct-Control systems tend to benefit the most from the knowledge gained from additional information and hands-on experimentation.
- Remote-Control systems tend to benefit somewhat from the knowledge gained from additional information and hands-on experimentation.
- Autonomous systems tend to benefit the least from the knowledge gained from additional information and hands-on experimentation, and in fact, the additional work performed contributes added insights to only 18% of the cases.

> **"The formation of trust in technology is governed by two constructs: reason-based trust and experience-based trust.**

The Three Systems Are Statistically Different with No Intervening Variables

The research questions we are attempting to answer in this section are:

- *Will different users of the technology with different backgrounds affect the results? That is, are there any controllable or blocking variables that need additional attention?*

Using the ANOVA with Blocking Variables model, we see the results in Table 8. In the experiment, the active-duty military either had experience with similar technology or they did not. The ANOVA test is run with blocking or controlling the user background.

TABLE 8. ANOVA WITH RANDOMIZED BLOCKS

Model Inputs:
VAR296; VAR297; VAR298
SUS(A), SUS(B), SUS(C)

ANOVA Randomized Blocks Multiple Treatments

	DF	SS	MS	F Stat	*P*-value
Block Factor (Row)	18	4384.65	243.59	1.5282	0.1367
Treatment Factor (Column)	2	11369.96	5684.98	5.6650	**0.0000**
Error	36	5738.38	159.40		
Total	56	21492.98			
F Critical (Treatment) @ 0.01	5.247893				
F Critical (Blocking) @ 0.01	2.479730				

Note. SUS = System Usability Score (for systems A, B, and C).

We conclude that:

- The treatment factor indicates that statistically significantly different results are shown among the three systems, but whether a soldier has experience with similar technology does not affect the results.

Nonparametric Kruskal–Wallis

The research question we are attempting to answer in this section is:

- *Does a nonparametric approach yield different results than a parametric model?*

Table 9 shows the results from the nonparametric Kruskal–Wallis test. As discussed, this test is the nonparametric equivalence of the ANOVA. Researchers use it to confirm the results of the ANOVA.

TABLE 9. ANOVA AND KRUSKAL-WALLIS I

VARIABLES TESTED	ANOVA	K-W
VAR20; VAR31; VAR42	0.0008	0.0008
VAR21; VAR32; VAR43	0.0903	0.0116
VAR120; VAR131; VAR142	0.0264	0.0057
VAR124; VAR135; VAR146	0.0362	0.0317
VAR229; VAR240; VAR251	0.0000	0.0000
VAR230; VAR241; VAR252	0.0000	0.0000
VAR231; VAR242; VAR253	0.0000	0.0000
VAR232; VAR243; VAR254	0.0002	0.0000
VAR233; VAR244; VAR255	0.0601	0.0851
VAR237; VAR248; VAR259	0.0004	0.0000
VAR238; VAR249; VAR260	0.0000	0.0000
VAR239; VAR250; VAR261	0.0000	0.0248
VAR266; VAR276; VAR286	0.0000	0.0000
VAR267; VAR277; VAR287	0.0285	0.0239
VAR268; VAR278; VAR288	0.0003	0.0022
VAR269; VAR279; VAR289	0.0020	0.0162
VAR270; VAR280; VAR290	0.0000	0.0000
VAR271; VAR281; VAR291	0.0000	0.0000
VAR272; VAR282; VAR292	0.0351	0.0208
VAR273; VAR283; VAR293	0.0002	0.0007
VAR274; VAR284; VAR294	0.0000	0.0000

We conclude that:

- Comparable to the ANOVA (from Table 6), the Kruskal–Wallis results show that out of the 43 models, the same 21 combinations have statistical significance.

Table 10 shows the additional results from the nonparametric Kruskal–Wallis test. Similar to the ANOVA, the Kruskal–Wallis shows that out of the 33 models, the same 16 combinations show statistical significance.

TABLE 10. ANOVA AND KRUSKAL–WALLIS II

	ANOVA	KW
ANOVA & KW on VAR15; VAR115; VAR415	0.0000	0.0000
ANOVA & KW on VAR16; VAR116; VAR416	0.0000	0.0000
ANOVA & KW on VAR17; VAR117; VAR417	0.0000	0.0000
ANOVA & KW on VAR18; VAR118; VAR418	0.0000	0.0000
ANOVA & KW on VAR19; VAR119; VAR419	0.0008	0.0015
ANOVA & KW on VAR20; VAR120; VAR420	0.0003	0.0000
ANOVA & KW on VAR21; VAR121; VAR421	0.0001	0.0003
ANOVA & KW on VAR22; VAR122; VAR422	0.0000	0.0000
ANOVA & KW on VAR23; VAR123; VAR423	0.0001	0.0000
ANOVA & KW on VAR24; VAR124; VAR424	0.0000	0.0000
ANOVA & KW on VAR28; VAR128; VAR428	0.0232	0.0128
ANOVA & KW on VAR29; VAR129; VAR429	0.0157	0.0127
ANOVA & KW on VAR31; VAR131; VAR431	0.0114	0.0085
ANOVA & KW on VAR35; VAR135; VAR435	0.0089	0.0008
ANOVA & KW on VAR38; VAR138; VAR438	0.0472	0.0631
ANOVA & KW on VAR43; VAR143; VAR443	0.0324	0.0614

The Data Are Reliable and Valid

The research question we are attempting to answer in this section is:

- *Are the collected data reliable and valid for the research?*

The Interrater Reliability Test with Interclass Correlation (ICC) tests were run to determine if the data received were statistically reliable (Table 11). As mentioned, the ICC tests the reliability of the users' ratings by comparing the variability of all the ratings of the same subject to the total variation across all ratings and all users simultaneously. A high ICC indicates a high level of reliability (Mun, 2018).

TABLE 11. ICC AND RELIABILITY ANALYSIS		
Intercorrelation ICC Reliability Measures (ICC)		
Pre-Experiment Less Info	**ICC**	***P*-value**
A1:: VAR1:VAR14	0.3544	0.0000
A2:: VAR15:VAR25	0.2886	0.0000
A3:: VAR26:VAR36	0.2302	0.0000
A4:: VAR37:VAR47	0.2692	0.0000
A5:: VAR48:VAR51		
Pre-Experiment More Info	**ICC**	***P*-value**
B1:: VAR101:VAR114	0.3207	0.0000
B2:: VAR115:VAR125	0.2568	0.0000
B3:: VAR126:VAR136	0.2528	0.0000
B4:: VAR137:VAR147	0.2975	0.0000
B5:: VAR148:VAR151	0.1581	0.0016
Post Experiment	**ICC**	***P*-value**
VAR201:VAR214	0.5067	0.0000
VAR215:VAR228	0.4584	0.0000
VAR229:VAR239	0.6709	0.0000
VAR240:VAR250	0.2593	0.0000
VAR251:VAR261	0.2200	0.0000
VAR262:VAR265	0.3146	0.0000
VAR266:VAR275	0.6925	0.0000
VAR276:VAR285	0.2264	0.0000
VAR286:VAR295	0.2328	0.0000
VAR296:VAR298	0.6081	0.0000

We conclude that:

- The data show statistical significance, and we conclude that the collected data are reliable and valid for the research.
- The ICC ranges from 0.1581 to 0.3544 for the Pre-experiment stage for both Less Info and More Info, compared to a range from 0.2200 to 0.6925 for the Post-experiment results. In other words, the more hands-on experimentation, the higher the validity of the collected data.

The Systems Are Independent and Uncorrelated

The research question we are attempting to answer in this section is:

- *Are the three systems somehow correlated in terms of their value to the warfighter?*

Table 12 shows a sampling of the results from the linear and nonlinear correlation matrices.

TABLE 12. LINEAR AND NONLINEAR CORRELATION MATRIX			
Linear Correlation			
	VAR296	**VAR297**	**VAR298**
VAR296	1.000000	0.234553	0.279342
VAR297	0.234553	1.000000	0.065035
VAR298	0.279342	0.065035	1.000000
Linear Correlation *p*-Value			
VAR296	0.000000	0.333765	0.246782
VAR297	0.333765	0.000000	0.791381
VAR298	0.246782	0.791381	0.000000
Nonlinear Correlation			
VAR296	1.000000	0.206909	0.265491
VAR297	0.206909	1.000000	0.090518
VAR298	0.265491	0.090518	1.000000

We conclude that:

- It seems that very little correlation exists among the three final scores of the systems.

The results and conclusion make sense, as there should be very little relationship among the Direct-Control, Remote-Control, and Autonomous systems, especially when they are tested independently and at different times.

Each Level of Experimentation Yields Valuable Actionable Intelligence

The research questions we are attempting to answer in this section are:

- *Within each experimentation stage, are the three systems perceived to be different (Direct-Control vs. Remote-Control, Direct-Control vs. Autonomous, and Remote-Control vs. Autonomous systems)?*

- *Between the three levels of experimentation (Less Info, More Info, Live Experiments), are each of the subsections of the technology considered similar or different?*

Tables 13 and 14 show a summary of the results from the relevant *T*-tests and Mann–Whitney (MW) tests. Table 13 shows the results that answer the first question above whereas Table 14 answers the second research question above.

TABLE 13. PARAMETRIC T-TEST AND NONPARAMETRIC MANN–WHITNEY TEST I

Direct vs. Remote	T-Test *P*-value	MW *P*-value	Direct vs. Autonomous	T-Test *P*-value	MW *P*-value	Remote vs. Autonomous	T-Test *P*-value	MW *P*-value
VAR20; VAR31	0.4306	0.4388	VAR20; VAR42	**0.0008**	**0.0013**	VAR31; VAR42	**0.0011**	**0.0016**
VAR21; VAR32	0.1783	0.2914	VAR21; VAR43	**0.0146**	**0.0199**	VAR32; VAR43	0.1100	0.0782
VAR120; VAR131	0.1728	0.1584	VAR120; VAR142	**0.0025**	**0.0043**	VAR131; VAR142	**0.0442**	0.0684
VAR120; VAR131	**0.0069**	**0.0180**	VAR124; VAR146	**0.0496**	0.0902	VAR135; VAR146	0.1984	0.2030
VAR229; VAR240	**0.0008**	**0.0011**	VAR229; VAR251	**0.0000**	**0.0000**	VAR240; VAR251	**0.0469**	**0.0362**
VAR230; VAR241	**0.0000**	**0.0000**	VAR230; VAR252	**0.0000**	**0.0000**	VAR241; VAR252	0.3202	0.3413
VAR231; VAR242	**0.0000**	**0.0000**	VAR231; VAR253	**0.0000**	**0.0000**	VAR242; VAR253	0.0577	0.0626
VAR232; VAR243	**0.0000**	**0.0002**	VAR232; VAR254	**0.0000**	**0.0001**	VAR243; VAR254	0.1160	0.1336
VAR233; VAR244	0.1970	0.3795	VAR233; VAR255	**0.0064**	**0.0148**	VAR244; VAR255	0.0847	0.0722
VAR237; VAR248	**0.0211**	**0.0068**	VAR237; VAR259	**0.0000**	**0.0001**	VAR248; VAR259	**0.0207**	**0.0178**
VAR238; VAR249	**0.0010**	**0.0006**	VAR238; VAR260	**0.0000**	**0.0000**	VAR249; VAR260	**0.0126**	**0.0212**
VAR239; VAR250	0.3377	0.2651	VAR239; VAR261	0.4549	0.3521	VAR250; VAR261	0.3706	0.3851
VAR266; VAR276	**0.0012**	**0.0027**	VAR266; VAR286	**0.0000**	**0.0000**	VAR276; VAR286	**0.0007**	**0.0016**
VAR267; VAR277	**0.0461**	0.0994	VAR267; VAR287	**0.0015**	**0.0044**	VAR277; VAR287	0.1865	0.2195
VAR268; VAR278	0.1402	0.2110	VAR268; VAR288	**0.0001**	**0.0010**	VAR278; VAR288	**0.0037**	**0.0090**
VAR269; VAR279	0.3237	0.2919	VAR269; VAR289	**0.0006**	**0.0015**	VAR279; VAR289	**0.0036**	**0.0043**
VAR270; VAR280	**0.0007**	**0.0026**	VAR270; VAR290	**0.0000**	**0.0000**	VAR280; VAR290	0.1646	0.2060
VAR271; VAR281	**0.0000**	**0.0002**	VAR271; VAR291	**0.0000**	**0.0000**	VAR281; VAR291	**0.0008**	**0.0019**
VAR272; VAR282	0.3549	0.4883	VAR272; VAR292	**0.0070**	**0.0128**	VAR282; VAR292	**0.0305**	**0.0191**
VAR273; VAR283	0.4451	0.4362	VAR273; VAR293	**0.0001**	**0.0004**	VAR283; VAR293	**0.0001**	**0.0005**
VAR274; VAR284	**0.0048**	**0.0080**	VAR274; VAR294	**0.0000**	**0.0000**	VAR284; VAR294	**0.0034**	**0.0040**

TABLE 14. PARAMETRIC T-TEST AND NONPARAMETRIC MANN-WHITNEY TEST II

Less Info vs. More Info	T-Test *P*-value	MW *P*-value	Less Info vs. Live Experiment	T-Test *P*-value	MW *P*-value	More Info vs. Live Experiment	T-Test *P*-value	MW *P*-value
VAR15; VAR115	0.2858	0.2593	VAR15; VAR415	**0.0000**	**0.0000**	VAR115; VAR415	**0.0000**	**0.0001**
VAR16; VAR116	0.3564	0.4038	VAR16; VAR416	**0.0000**	**0.0000**	VAR116; VAR416	**0.0000**	**0.0000**
VAR17; VAR117	0.2419	0.2438	VAR17; VAR417	**0.0000**	**0.0000**	VAR117; VAR417	**0.0000**	**0.0000**
VAR18; VAR118	0.2298	0.4038	VAR18; VAR418	**0.0000**	**0.0000**	VAR118; VAR418	**0.0000**	**0.0000**
VAR19; VAR119	0.0968	0.1094	VAR19; VAR419	**0.0037**	**0.0043**	VAR119; VAR419	**0.0005**	**0.0010**
VAR20; VAR120	0.4280	0.1704	VAR20; VAR420	**0.0000**	**0.0004**	VAR120; VAR420	**0.0000**	**0.0010**
VAR21; VAR121	0.3632	0.3711	VAR21; VAR421	**0.0000**	**0.0002**	VAR121; VAR421	**0.0000**	**0.0005**
VAR22; VAR122	0.4853	0.4227	VAR22; VAR422	**0.0000**	**0.0000**	VAR122; VAR422	**0.0000**	**0.0000**
VAR23; VAR123	0.1156	**0.0489**	VAR23; VAR423	**0.0012**	**0.0003**	VAR123; VAR423	**0.0000**	**0.0000**
VAR24; VAR124	0.0518	0.0610	VAR24; VAR424	**0.0000**	**0.0000**	VAR124; VAR424	**0.0000**	**0.0000**
VAR28; VAR128	**0.0148**	**0.0388**	VAR28; VAR428	**0.0078**	**0.0055**	VAR128; VAR428	**0.0271**	0.1005
VAR29; VAR129	**0.0102**	**0.0192**	VAR29; VAR429	**0.0059**	**0.0022**	VAR129; VAR429	0.1867	0.0515
VAR31; VAR131	0.1438	0.1420	VAR31; VAR431	**0.0016**	**0.0017**	VAR131; VAR431	**0.0362**	**0.0254**
VAR35; VAR135	**0.0383**	0.0771	VAR35; VAR435	**0.0032**	**0.0016**	VAR135; VAR435	0.0617	**0.0142**
VAR38; VAR138	0.0986	0.0883	VAR38; VAR438	**0.0130**	**0.0395**	VAR138; VAR438	0.0948	0.1769
VAR43; VAR143	0.1612	0.1650	VAR43; VAR443	**0.0085**	**0.0123**	VAR143; VAR443	**0.0484**	**0.0486**

We conclude that:

- Within each experimentation stage, the three systems are indeed perceived to be different.
  - Direct-Control vs. Autonomous shows the most amount of difference, regardless of the experimental stage.
  - A majority of the Direct-Control vs. Remote Control and Remote-Control vs. Autonomous systems also showed differences, although less than the Direct-Control vs. Autonomous systems.
- Between the three levels of experimentation (Less Info, More Info, live experiments), each subsection of the technology is considered statistically different.

- ◦ Live experimentation shows a significant difference in the information and knowledge gathered.
- ◦ Live experimentation can be concluded to have significant value and insight.
- ◦ The difference between Less Info and More Info without any hands-on experimentation is only limited. In other words, having additional information on paper, without the ability to perform hands-on experimentation, yields little difference and only minor benefits.

Predictability Without Experimentation Is Very Limited

The research question we are attempting to answer in this section is:

- *Can the final outcome of a detailed experiment be predicted by performing some basic Pre-experimental survey?*

If the research question above is found to be predictable, this would save the DoD considerable time and expense. Results from detailed experimentation can be predicted from basic preliminary review of the technology.

Table 15 shows a sampling of the results from a multivariate regression model. Little to no statistical significance is discernible when using Pre-experimental data to predict the outcomes of the Post-experiment scores.

Multiple linear and nonlinear interacting multivariate regressions were also run, and none seems to exhibit coefficients of determination greater than 50% and adjusted coefficients of determination greater than 25%.

TABLE 15. LIMITED PREDICTABILITY WITH LINEAR AND NONLINEAR MULTIVARIATE REGRESSION

Model Inputs:
VAR296 vs. VAR15; VAR16; VAR17; VAR18; VAR19; VAR20; VAR21; VAR22; VAR23; VAR24; VAR25
SUSA vs. PU1, PU2, PU3, PU4, PEOU1, PEOU2, PEOU3, PEOU4, IU1, IU2, IU3

Multiple R	0.85341	Maximum Log Likelihood	-52.79311
R-Square	0.72830	Akaike Info Criterion (AIC)	6.82033
Adjusted R-Square	0.30135	Bayes Schwarz Criterion (BSC)	7.41682
Standard Error	7.28268	Hannan-Quinn Criterion (HQC)	6.92128

	Coeff	Std. Error	T-stat	*P*-value	Lower 5%	Upper 95%
Intercept	135.82272	26.21155	5.18179	0.00128	73.84226	197.80319
VAR X1	-1.78874	5.52795	-0.32358	0.75571	-14.86027	11.28279
VAR X2	0.02206	4.87473	0.00452	0.99652	-11.50484	11.54895
VAR X3	-13.67128	6.12796	-2.23097	0.06088	-28.16161	0.81904
VAR X4	-9.34621	6.28587	-1.48686	0.18065	-24.20993	5.51752
VAR X5	-1.40361	5.80732	-0.24170	0.81594	-15.13574	12.32853
VAR X6	-5.81092	3.63238	-1.59976	0.15369	-14.40012	2.77829
VAR X7	-2.34249	4.29174	-0.54581	0.60215	-12.49084	7.80587
VAR X8	1.71980	3.64092	0.47235	0.65105	-6.88960	10.32921
VAR X9	17.00398	5.38884	3.15541	0.01603	4.26140	29.74656
VAR X10	2.09003	3.88437	0.53806	0.60721	-7.09505	11.27512
VAR X11	-5.90165	2.22358	-2.65412	0.03275	-11.15957	-0.64372

ANOVA	DF	SS	MS	F	*p*-Value
Regression	11	995.19	90.47	1.70580	0.24525
Residual	7	371.26	53.04		
Total	18	1366.45			

Hypothesis Test
Critical F-statistic (99% confidence with DFR1 and DFR2) : 6.538166
Critical F-statistic (95% confidence with DFR1 and DFR2) : 3.603037
Critical F-statistic (90% confidence with DFR1 and DFR2) : 2.683924

Table 16 shows a principal component analysis and factor analysis result where the multiple variables were reduced further to see if there would be any improvements in the multivariate regression, but the results similarly indicate very low predictive power in the Pre-experiment results.

TABLE 16. PRINCIPAL COMPONENT ANALYSIS

Model Inputs:
VAR23:VAR33
PU1, PU2, PU3, PU4, PEOU1, PEOU2, PEOU3, PEOU4, IU1, IU2, IU3
* indicates negative values

Cum Proportions:										
55.05%	75.51%	85.26%	90.59%	94.76%	96.74%	97.92%	98.97%	99.57%	99.87%	100.00%
Eigenvectors:										
0.3537	*0.2475	*0.1379	*0.0953	0.1383	*0.3637	*0.0508	*0.4055	0.5314	*0.0935	*0.4195
0.3592	*0.2186	0.0260	0.1763	0.1098	*0.1431	*0.7667	0.1402	*0.1323	0.2344	0.2811
Eigenvalues (Arranged and Ranked):										
6.0552	2.2509	1.0725	0.5861	0.4586	0.2184	0.1292	0.1157	0.0666	0.0320	0.0148

A traditional ordinary least squares multivariate regression also does not make too much sense in that no one-to-one correspondence is detected among the data rows. That is, different active-duty military from the same unit participated in the three experimental stages. This means that the responses of one soldier will not correspond to the same perception of another soldier testing another system during a different stage. This explains partly the low predictability of Post-experiment results using Pre-experiment data.

> **Having additional information on paper, without the ability to perform hands-on experimentation, yields little difference and only minor benefits.**

Additional sophisticated methods were performed, such as bootstrapping the regression, where an empirical bootstrap of the data was nonparametrically simulated and bootstrapped, then regression models were run. The process was repeated thousands of times. Figures 3, 4, and Table 17 illustrate the results. Only 9% to 12% of the time will a single variable be considered statistically significant, and the goodness-of-fit predictability levels vary widely, from 18% to 95%, depending on the specific issue under study. No consistent and valid predictive power is apparent in the Pre-experiment data. This concurs with the two-variable *T*-tests and MW tests shown previously where we do see significant and valuable insights exist when hands-on experimentation is performed, which means without these experiments, paper-based cursory system knowledge is insufficient to identify the true value and risks of a system.

TABLE 17. BOOTSTRAP REGRESSION III

Variable	IU1	IU2	IU3	PEOU1	PEOU2	PEOU3	PEOU4	PU1	PU2	PU3	PU4	R Square
Number of Datapoints	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
Mean	0.5035	0.4958	0.4816	0.5066	0.4952	0.5217	0.5016	0.5064	0.5018	0.4944	0.5035	**0.6150**
Median	0.5132	0.4926	0.4652	0.5103	0.5026	0.5277	0.4935	0.5071	0.4971	0.4974	0.5057	**0.6226**
Standard Deviation	0.2886	0.2922	0.2931	0.2853	0.2889	0.2893	0.2867	0.2801	0.2862	0.2846	0.2917	0.1553
Variance	0.0833	0.0854	0.0859	0.0814	0.0840	0.0837	0.0822	0.0785	0.0819	0.0810	0.0851	2.41%
Coefficient of Variation	57.32%	58.94%	60.86%	56.32%	58.54%	55.46%	57.17%	55.32%	57.04%	57.57%	57.93%	0.2525
Maximum	1.0000	0.9974	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9712
Minimum	0.0021	0.0004	0.0004	0.0009	0.0005	0.0051	0.0003	0.0020	0.0009	0.0009	0.0005	0.1263
Range	0.9979	0.9970	0.9996	0.9991	0.9995	0.9949	0.9997	0.9980	0.9991	0.9991	0.9995	0.8449
Skewness	-0.0422	0.0362	0.0921	-0.0761	-0.0067	-0.0855	0.0071	0.0015	0.0005	0.0199	-0.0284	-0.2432
Kurtosis	-1.2087	-1.2455	-1.2177	-1.1780	-1.2168	-1.2010	-1.1436	-1.1199	-1.1846	-1.1535	-1.2117	-0.2463
25% Percentile	0.2408	0.2426	0.2211	0.2619	0.2352	0.2718	0.2533	0.2696	0.2630	0.2519	0.2550	0.5091
75% Percentile	0.7496	0.7566	0.7372	0.7596	0.7483	0.7652	0.7395	0.7330	0.7448	0.7325	0.7575	72.50%
Error Precision at 95%	3.56%	3.66%	3.78%	3.50%	3.63%	3.44%	3.55%	3.43%	3.54%	3.57%	3.60%	0.0157
5% Percentile	0.0463	0.0547	0.0409	0.0463	0.0443	0.0653	0.0459	0.0568	0.0501	0.0465	0.0430	0.3515
10% Percentile	0.1017	0.0985	0.0854	0.0986	0.0931	0.1071	0.1045	0.1205	0.1030	0.0985	0.0906	0.4041
20% Percentile	0.1984	0.1979	0.1840	0.2091	0.1917	0.2191	0.2021	0.2212	0.2137	0.2060	0.1933	0.4786
30% Percentile	0.3000	0.2837	0.2786	0.3223	0.2888	0.3305	0.3096	0.3198	0.3107	0.2929	0.2997	0.5361
40% Percentile	0.4165	0.3776	0.3701	0.4166	0.3913	0.4321	0.4187	0.4094	0.3993	0.3930	0.4010	0.5781
50% Percentile	0.5129	0.4914	0.4645	0.5098	0.5024	0.5260	0.4929	0.5062	0.4946	0.4962	0.5043	0.6224
60% Percentile	0.6109	0.5961	0.5733	0.6227	0.5969	0.6353	0.5875	0.6018	0.6012	0.5886	0.6142	0.6661
70% Percentile	0.7032	0.6981	0.6784	0.7089	0.6980	0.7279	0.6947	0.6931	0.7002	0.6800	0.7015	0.7013
80% Percentile	0.7940	0.8051	0.7959	0.7999	0.7913	0.8209	0.7956	0.7899	0.7930	0.7898	0.8053	0.7545
90% Percentile	0.8957	0.9014	0.8895	0.8876	0.8964	0.9119	0.9032	0.8988	0.9020	0.8868	0.9035	0.8156
95% Percentile	0.9418	0.9455	0.9498	0.9368	0.9448	0.9543	0.9536	0.9498	0.9419	0.9447	0.9570	0.8594
99% Percentile	0.9915	0.9894	0.9920	0.9912	0.9875	0.9914	0.9905	0.9923	0.9955	0.9914	0.9904	0.9344
Certainty Value 0.01	0.80%	1.40%	0.80%	1.10%	1.10%	0.40%	1.40%	1.20%	0.80%	1.30%	1.00%	
Certainty Value 0.05	5.52%	4.71%	6.32%	5.02%	5.52%	3.81%	5.22%	4.21%	4.81%	5.02%	5.52%	
Certainty Value 0.1	9.83%	10.13%	11.94%	10.03%	10.73%	9.33%	9.53%	7.92%	9.83%	10.23%	10.93%	

The main conclusion from the analysis is:

The final detailed experimental results cannot be sufficiently predicted by using Pre-experiment survey data, regardless of how much nonexperimental, paper-based information is provided to the user.

FIGURE 3. BOOTSTRAP REGRESSION I

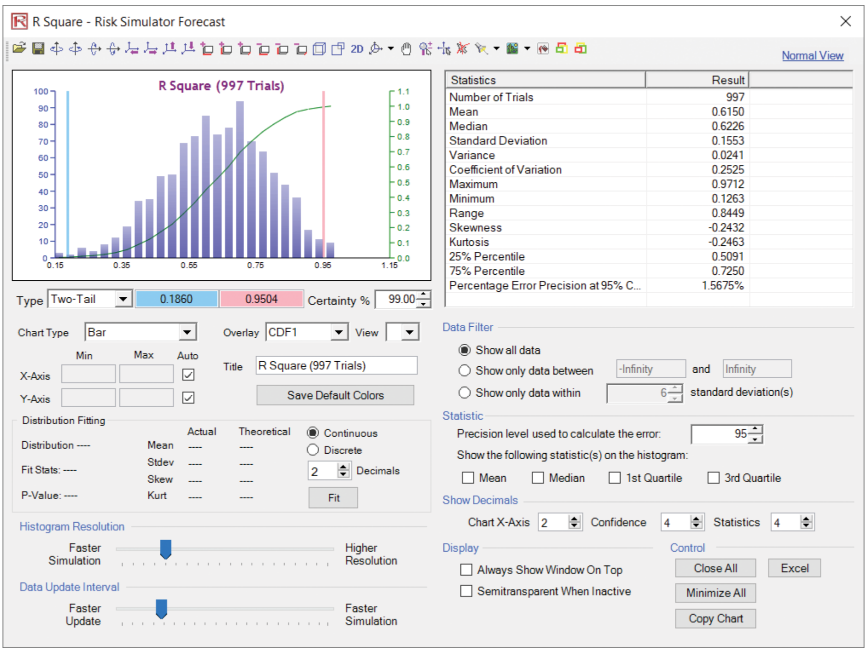
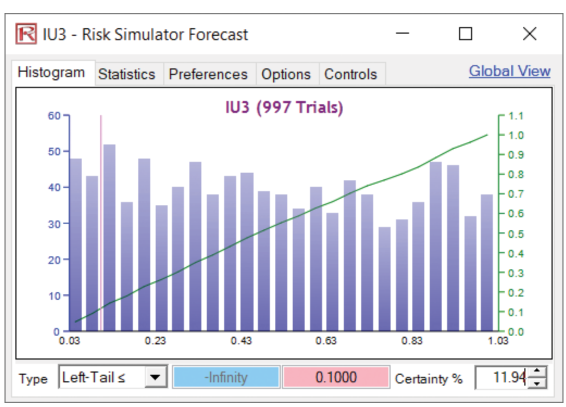


FIGURE 4. BOOTSTRAP REGRESSION II



Limitations of Research

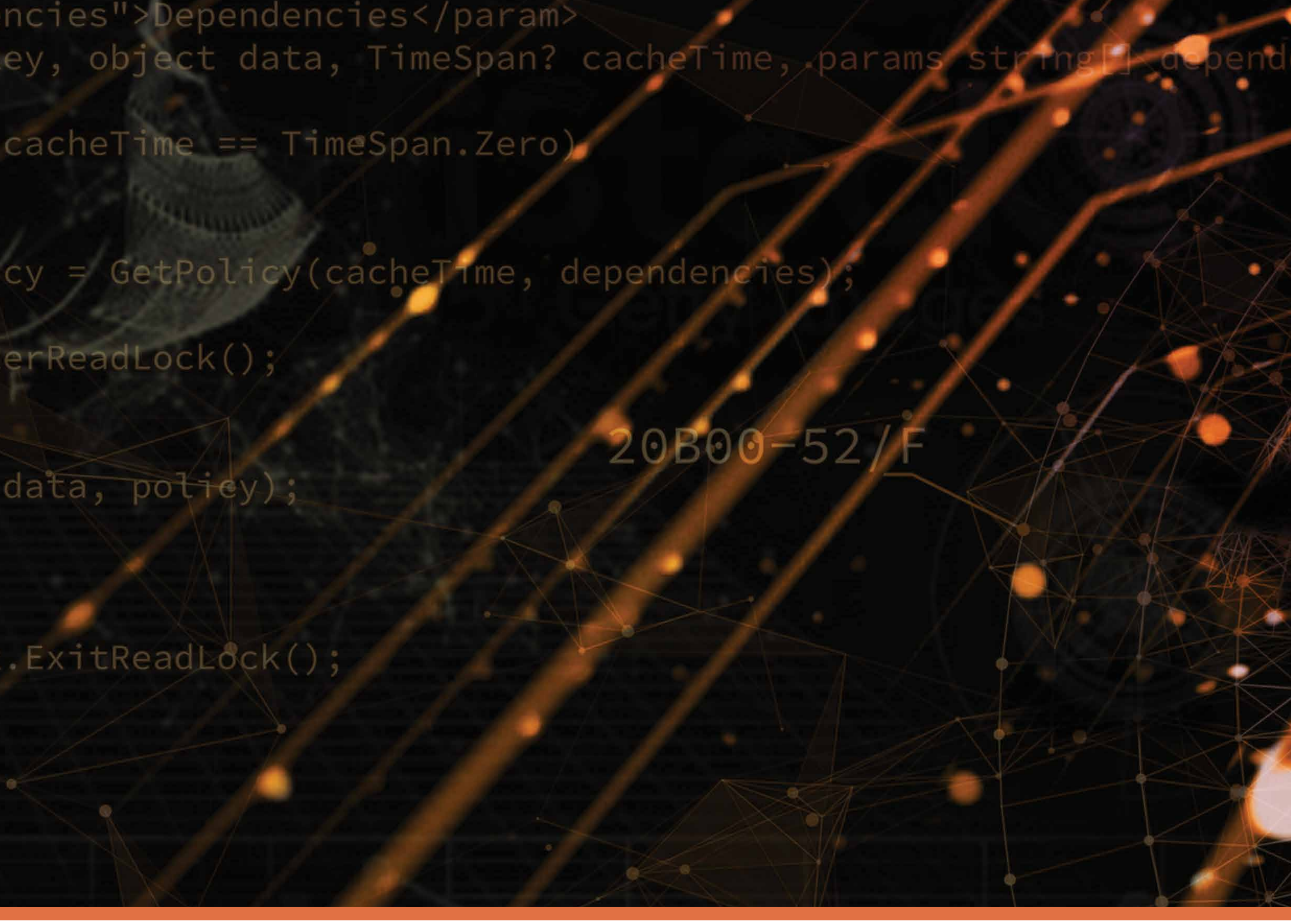
The investigators used secondary data collected by the U.S. Department of Defense. The sample size was not within the control of the investigators and represented a smaller than desired number of participants.

Conclusions

The topic of trust in technology is increasingly important to the DoD as outlined in the Defense Science Board Study on Autonomy (David & Nielsen, 2016), which states, "There is a need to build trust in autonomous systems while also improving the trustworthiness of autonomous capabilities. These are enablers that align RDT&E [research, development, test & evaluation] processes to more rapidly deliver autonomous capabilities to DoD missions."

This work involves the introduction of novel ideas to existing theories that relate to the formation of trust. This research focuses on the impact of trust towards the adoption of autonomous systems. We have established that trust involves a user assuming some level of risk. The only literature available on technology trust involves situations that expose users to insignificant levels of risk. We posit that our research conducted on technology used in high-risk military application will reveal causality not identified in previous trust research.

This research tests theories of anthropomorphism and system hierarchy by manipulating the amount of information to observe the impact on the formation of initial, reason-based, technology trust. The article begins to answer the question of whether or not it is possible to predict and potentially capture trust in technology used for high-risk military applications. If a causal relationship exists between technology features and acceptance, it could greatly reduce the time and expense of adopting new technologies. The initial findings of this research indicate that manipulating familiarization with technology through the use of anthropomorphic categories, without the use of experience-based data or the ability to perform hands-on experimentation, yields little difference and only minor benefits. This article warrants further research to identify the influence of experience-based trust on the formation of reason-based trust.



References

Adams, B. D., & Webb, R. D. (2002, September 16–20). Trust in small military teams. *In Proceedings of the 7th International Command and Control Research and Technology Symposium* (pp. 1-20), Québec City, Canada. http://dodccrp.org/events/7th_ICCRTS/Tracks/pdf/006.PDF

Castelfranchi, C., & Falcone, R. (2010). *Trust theory: A socio-cognitive and computational model.* Wiley Online Library. https://onlinelibrary.wiley.com/doi/book/10.1002/9780470519851

Cho, J.-H., Chan, K., & Adali, S. (2015). *A survey on trust modeling. ACM Computing Surveys (CSUR), 48*(2), 28. https://doi.org/10.1145/2815595

David, R. A., & Nielsen, P. (2016). *Defense science board summer study on autonomy.* Defense Science Board. https://dsb.cto.mil/reports/2010s/DSBSS15.pdf

Gefen, D., Karahanna, E., & Straub, D. W. (2003). Trust and TAM in online shopping: An integrated model. *MIS Quarterly, 27*(1), 51–90. https://www.jstor.org/stable/30036519?seq=1

McKnight, D. H., Carter, M., Thatcher, J. B., & Clay, P. F. (2011). Trust in a specific technology: An investigation of its components and measures. *ACM Transactions on Management Information Systems (TMIS), 2*(2), 12. https://dl.acm.org/doi/10.1145/1985347.1985353

Mun, J. (2018). *Quantitative research methods.* Thompson-Shore & IIPER Press. https://www.amazon.com/gp/product/B083Q2MVBJ/

Pak, R., Fink, N., Price, M., Bass, B., & Sturre, L. (2012). Decision support aids with anthropomorphic characteristics influence trust and performance in younger and older adults. *Ergonomics, 55*(9), 1059–1072. https://psycnet.apa.org/record/2012-22389-007P

Rempel, J. K., Holmes, J. G., & Zanna, M. P. (1985). Trust in close relationships. *Journal of Personality and Social Psychology, 49*(1), 95–112. https://psycnet.apa.org/record/1985-30794-001

Schaefer, K. E., Chen, J. Y. C., Szalma, J. L., & Hancock, P. A. (2016). A meta-analysis of factors influencing the development of trust in automation: Implications for understanding autonomy in future systems. *Human Factors, 20*(10), 1–24. https://www.researchgate.net/publication/299371857_A_Meta-Analysis_of_Factors_Influencing_the_Development_of_Trust_in_Automation_Implications_for_Understanding_Autonomy_in_Future_Systems

Tétard, F., & Collan, M. (2009, January 5–8). Lazy user theory: A dynamic model to understand user selection of products and services. In Ralph H. Sprague, Jr. (Chair), *Proceedings of the 42nd Hawaii International Conference on System Sciences* (pp. 1–9), Waikoloa, Big Island, HI. https://ieeexplore.ieee.org/xpl/conhome/4755313/proceeding

Venkatesh, V., & Bala, H. (2008). Technology acceptance model 3 and a research agenda on interventions. *Decision Sciences, 39*(2), 273–315. https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-5915.2008.00192.x

Waytz, A., Heafner, J., & Epley, N. (2014). The mind in the machine: Anthropomorphism increases trust in an autonomous vehicle. *Journal of Experimental Social Psychology, 52,* 113–117. https://www.scholars.northwestern.edu/en/publications/the-mind-in-the-machine-anthropomorphism-increases-trust-in-an-au

APPENDIX A
Research Instrument

The investigators used secondary data collected by the U.S. Department of Defense. Data collection occurred in two phases. Phase one of the data collection was conducted in a controlled and distraction-free classroom environment and involved completion of a user survey by two randomly selected groups of active-duty military from within a single unit tasked with a high-risk mission. Both groups participated in separate morning sessions lasting 1 hour each. The second session started immediately following completion of the first session. Each group was provided with identical overviews of a high-risk military scenario that would be completed by deploying three technology systems rather than human operators. The independent variable "system presentation" was manipulated between the first and second groups. The second independent variable, "system control" was provided to all participants in the form of three separate technologies.

Phase two of the experiment was conducted in the field and involved the hands-on testing of the three technologies introduced during the phase one survey. Phase two of the experiment was conducted 6 months after the classroom survey of phase one. A total of 15 participants were selected from the same military unit as in the phase one survey. This experiment was conducted over a 12-day period. The first 3 days were reserved for training, and the subsequent 9 days were used to test the operational capabilities of the systems in the high-risk scenario presented in phase one. The day after the field experimentation concluded, all participants gathered in a controlled classroom environment to respond to the same user survey provided in phase one.

The investigators used secondary data collected by the U.S. Department of Defense. That data collection activity was ruled not human subjects research by the governing Institutional Review Board (IRB) in accordance with Secretary of the Navy Instruction (SECNAVINST 3900.39E), December 19, 2017. The data to which the Naval Postgraduate School (NPS) investigators have access do not contain data that are personally identifiable. Therefore, the presented activity was deemed not human subjects research by NPS IRB.

APPENDIX B

Survey Questions

Survey			Pre-Experiment		Post-Experiment		
Question	**Category**	**Code**	**Less Info**	**More Info**	**Most Info**	**Usability Questions**	**Usability Scores**
Loss of system endurance (decreased operating time)	Risk of Failure	HA1	VAR1	VAR101	VAR401	Q1	VAR266
Loss of power (unable to overcome large obstacles)	Risk of Failure	HA2	VAR2	VAR102	VAR402	Q2	VAR266
Loss of agility (limited range of motion)	Risk of Failure	HA3	VAR3	VAR103	VAR403	Q3	VAR266
Loss of speed (operates slowly)	Risk of Failure	HA4	VAR4	VAR104	VAR404	Q4	VAR266
Only have direct control (radio/ autonomous have failed)	Risk of Failure	AL1	VAR5	VAR105	VAR405	Q5	VAR270
Only have radio control (direct/ autonomous have failed)	Risk of Failure	AL2	VAR6	VAR106	VAR406	Q6	VAR271
Only have autonomous operation (direct/radio have failed)	Risk of Failure	AL3	VAR7	VAR107	VAR407	Q7	VAR272
Loss of ability to store data (bad memory)	Risk of Failure	AL4	VAR8	VAR108	VAR408	Q8	VAR273
Slow response to commands (bad processor)	Risk of Failure	AL5	VAR9	VAR109	VAR409	Q9	VAR274
Loss of ability to obtain imagery (video)	Risk of Failure	LN1	VAR10	VAR110	VAR410	Q10	VAR275
Loss of ability to obtain environmental data	Risk of Failure	LN2	VAR11	VAR111	VAR411	Q1	VAR276
Loss of ability to geolocate/ navigate (GPS)	Risk of Failure	LN3	VAR12	VAR112	VAR412	Q2	VAR277
Loss of comms needed to send sensor data (no system transmit)	Risk of Failure	LN4	VAR13	VAR113	VAR413	Q3	VAR278
Loss of comms needed to control sensors (no system receive)	Risk of Failure	LN5	VAR14	VAR114	VAR414	Q4	VAR279
This system would improve my performance	Direct	PU1	VAR15	VAR115	VAR415	Q5	VAR280
The system would increase my accuracy	Direct	PU2	VAR16	VAR116	VAR416	Q6	VAR281
The system would enhance my effectiveness	Direct	PU13	VAR17	VAR117	VAR417	Q7	VAR282
Overall, this system would be useful	Direct	PU4	VAR18	VAR118	VAR418	Q8	VAR283
The operational use of this system is clear and understandable	Direct	PEOU1	VAR19	VAR119	VAR419	Q9	VAR284
Using this system should not require a lot of my mental effort	Direct	PEOU2	VAR20	VAR120	VAR420	Q10	VAR285
It should be easy to get this system to do what I want it to do	Direct	PEOU3	VAR21	VAR121	VAR421	Q1	VAR286
Overall, this system would be easy to use	Direct	PEOU4	VAR22	VAR122	VAR422	Q2	VAR287
Given the chance, I would use this system	Direct	IU1	VAR23	VAR123	VAR423	Q3	VAR288
It is likely that I would recommend this system	Direct	IU2	VAR24	VAR124	VAR424	Q4	VAR289
I have been exposed to this technology in the past	Direct	IU3	VAR25	VAR125	VAR425	Q5	VAR290
This system would improve my performance	Remote	PU1	VAR26	VAR126	VAR426	Q6	VAR291

Survey Questions (continued)

Survey			Pre-Experiment		Post-Experiment		
Question	**Category**	**Code**	**Less Info**	**More Info**	**Most Info**	**Usability Questions**	**Usability Scores**
The system would increase my accuracy	Remote	PU2	VAR27	VAR127	VAR427	Q7	VAR292
The system would enhance my effectiveness	Remote	PU3	VAR28	VAR128	VAR428	Q8	VAR293
Overall, this system would be useful	Remote	PU4	VAR29	VAR129	VAR429	Q9	VAR294
The operational use of this system is clear and understandable	Remote	PEOU1	VAR30	VAR130	VAR430	Q10	VAR295
Using this system should not require a lot of my mental effort	Remote	PEOU2	VAR31	VAR131	VAR431	SUSA	VAR296
It should be easy to get this system to do what I want it to do	Remote	PEOU3	VAR32	VAR132	VAR432	SUSB	VAR297
Overall, this system would be easy to use	Remote	PEOU4	VAR33	VAR133	VAR433	SUSC	VAR298
Given the chance, I would use this system	Remote	IU1	VAR34	VAR134	VAR434		
It is likely that I would recommend this system	Remote	IU2	VAR35	VAR135	VAR435		
I have been exposed to this technology in the past	Remote	IU3	VAR36	VAR136	VAR436		
This system would improve my performance	Autonomous	PU1	VAR37	VAR137	VAR437		
The system would increase my accuracy	Autonomous	PU2	VAR38	VAR138	VAR438		
The system would enhance my effectiveness	Autonomous	PU3	VAR39	VAR139	VAR439		
Overall, this system would be useful	Autonomous	PU4	VAR40	VAR140	VAR440		
The operational use of this system is clear and understandable	Autonomous	PEOU1	VAR41	VAR141	VAR441		
Using this system should not require a lot of my mental effort	Autonomous	PEOU2	VAR42	VAR142	VAR442		
It should be easy to get this system to do what I want it to do	Autonomous	PEOU3	VAR43	VAR143	VAR443		
Overall, this system would be easy to use	Autonomous	PEOU4	VAR44	VAR144	VAR444		
Given the chance, I would use this system	Autonomous	IU1	VAR45	VAR145	VAR445		
It is likely that I would recommend this system	Autonomous	IU2	VAR46	VAR146	VAR446		
I have been exposed to this technology in the past	Autonomous	IU3	VAR47	VAR147	VAR447		
Tasking is directly relevant to my job function	Control	C1	VAR48	VAR148	VAR262		
I am personally invested in learning how to conduct this mission	Control	C2	VAR49	VAR149	VAR263		
In general, I am comfortable learning how to use new technology	Control	C3	VAR50	VAR150	VAR264		
These technologies are critical for accomplishing this mission	Control	C4	VAR51	VAR151	VAR265		
What is your current job?	Demographic	D1	VAR52	VAR152			
How long in your current job?	Demographic	D2	VAR53	VAR153			
How long in the military?	Demographic	D3	VAR54	VAR154			

Author Biographies

Mr. Michael G. Anderson

is a PhD Candidate at the Naval Postgraduate School Information Sciences Department. His research interests include Technology Trust. Mr. Anderson has an extensive background in the design of wireless communication systems and the use and deployment of military technologies. His past experience includes over 10 years of work as adjunct faculty teaching university students and professional engineers and scientists in the areas of digital design and computer architecture. Mr. Anderson holds a BS and MS in Electrical Engineering from the University of Michigan, Ann Arbor.

(E-mail address: mganders@nps.edu)

Dr. Johnathan C. Mun

is a research professor at the Naval Postgraduate School and is a specialist in advanced decision analytics, quantitative risk modeling, strategic flexibility real options, predictive modeling, and portfolio optimization. He has authored 24 books; holds 22 patents and patents pending; created over a dozen software applications in advanced decision analytics; and has written over a hundred technical notes, journal articles, and white papers. He is currently the CEO of Real Options Valuation, Inc., and his prior positions include vice president of Analytics at Oracle/Crystal Ball and a senior manager at KPMG Consulting. Dr. Mun holds a PhD in Finance and Economics from Lehigh University, an MBA and MS from Southeastern University, and a BS in Physics and Biology from the University of Miami.

(E-mail address: jcmun@nps.edu)